

```
pwd
```

```
'C:\\Users\\Riya'
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv("C:\\Users\\Riya\\insurance.csv", index_col = 0)
```

```
df.head()
```

	sex	bmi	children	smoker	region	charges
age						
19	female	27.900	0	yes	southwest	16884.92400
18	male	33.770	1	no	southeast	1725.55230
28	male	33.000	3	no	southeast	4449.46200
33	male	22.705	0	no	northwest	21984.47061
32	male	28.880	0	no	northwest	3866.85520

Verified the data's shape to determine how many rows and columns there were.

```
df.shape
```

```
(1338, 6)
```

Checked the data types of the columns

```
df.dtypes
```

```
sex          object
bmi          float64
children     int64
smoker       object
region       object
charges      float64
dtype: object
```

Inspected the data for null values. Found how many null values there are in each column.

```
df.isnull().sum()
```

```
sex          0
bmi          0
children     0
smoker       0
region       0
charges      0
dtype: int64
```

```
# Found the list of columns present in the dataset
data = list(df.columns)
data

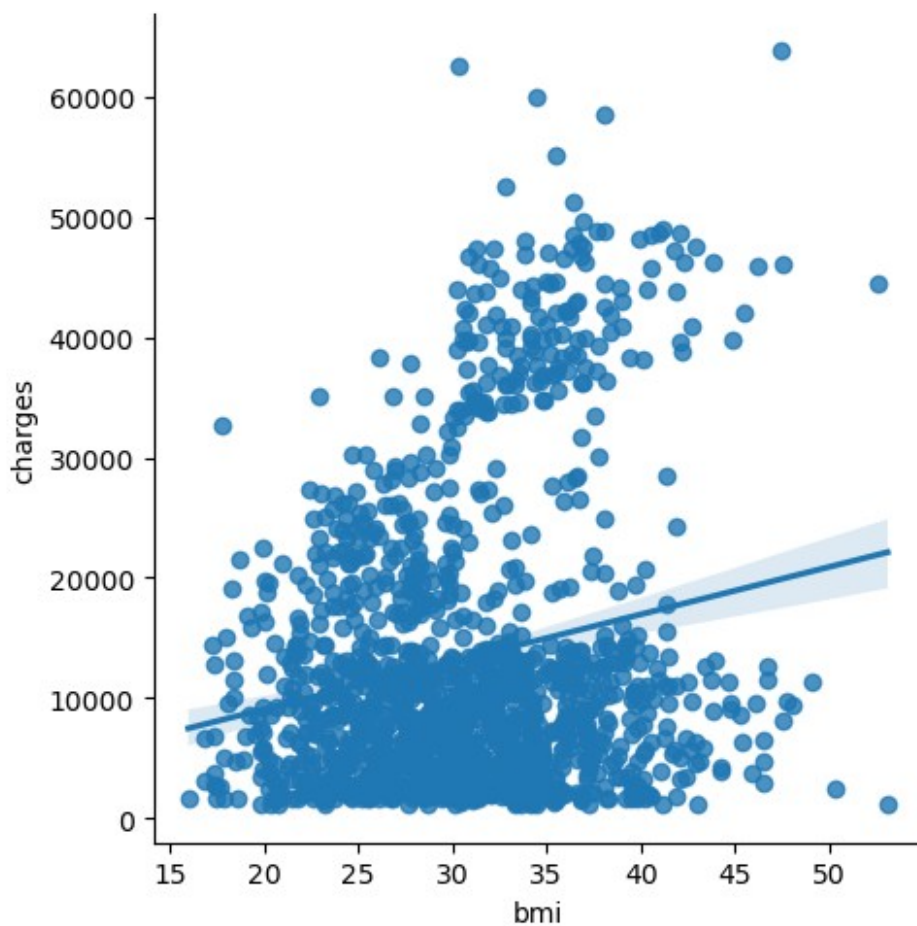
['sex', 'bmi', 'children', 'smoker', 'region', 'charges']

# Found the list of numeric columns present in the dataset
data = list(df._get_numeric_data().columns)
data

['bmi', 'children', 'charges']

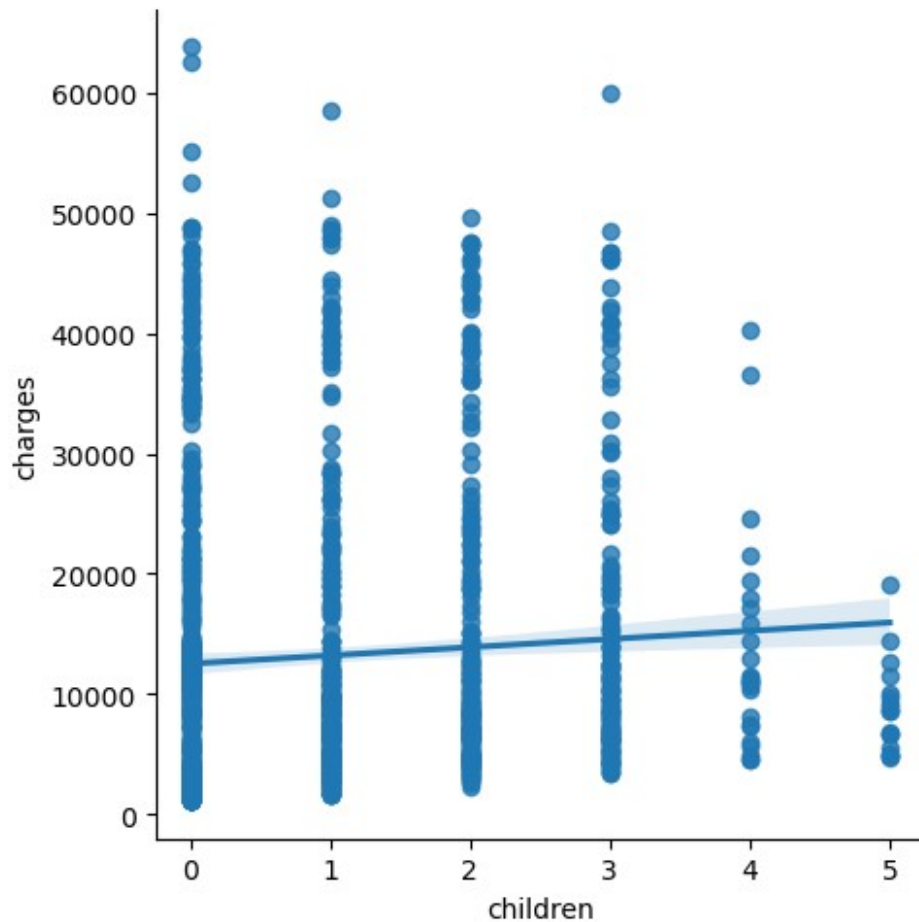
plt.figure(figsize=(7,4))
sns.lmplot(data = df,x='bmi',y='charges',legend=True)
plt.show()

<Figure size 700x400 with 0 Axes>
```

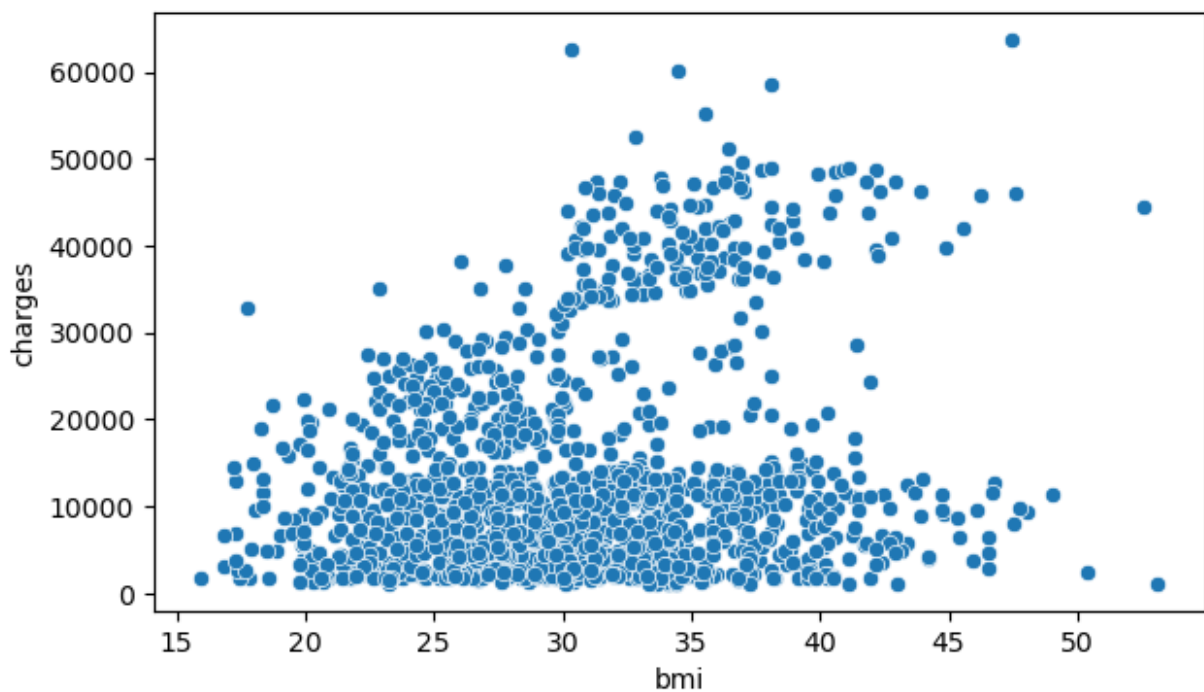


```
plt.figure(figsize=(7,4))
sns.lmplot(data = df,x='children',y='charges',legend=True)
plt.show()
```

<Figure size 700x400 with 0 Axes>

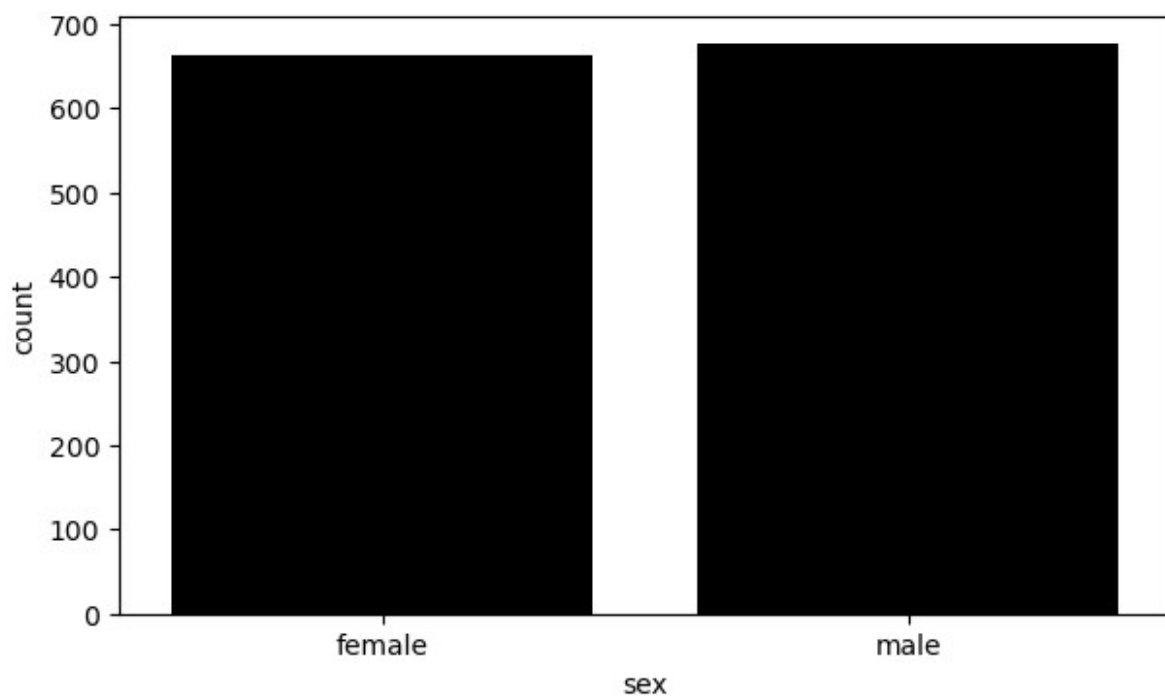


```
plt.figure(figsize=(7,4))
sns.scatterplot(x='bmi',y='charges',data = df)
plt.show()
```

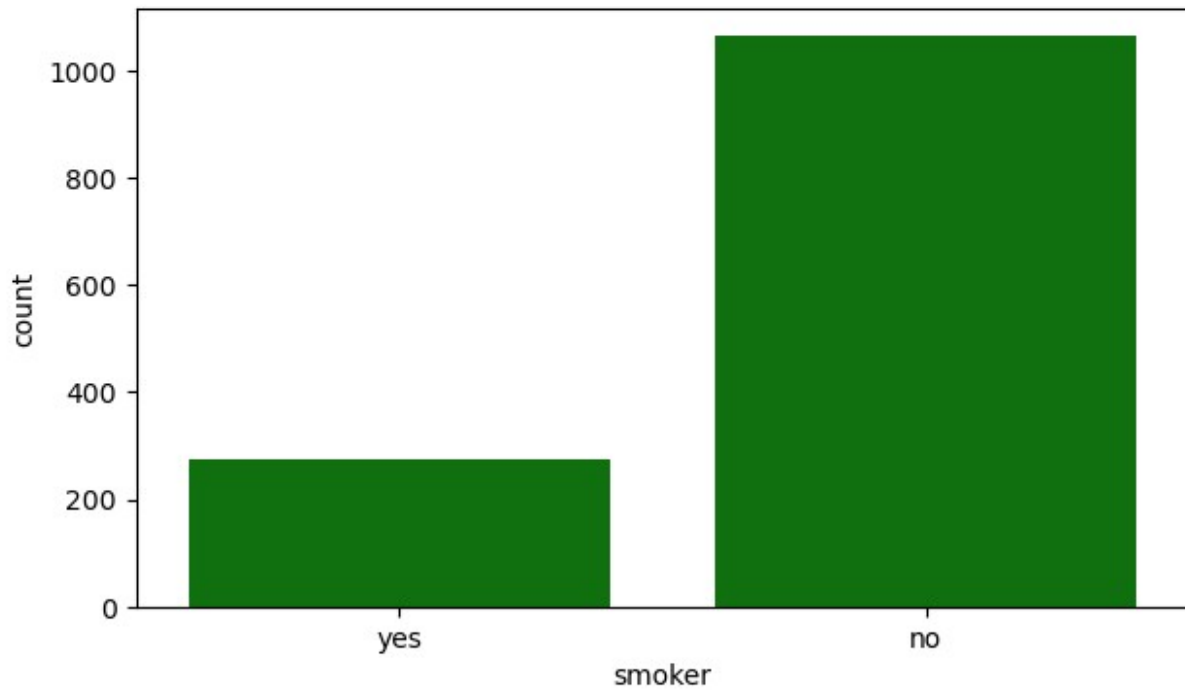


Created count plots of categorical variables to show number of observations in each category

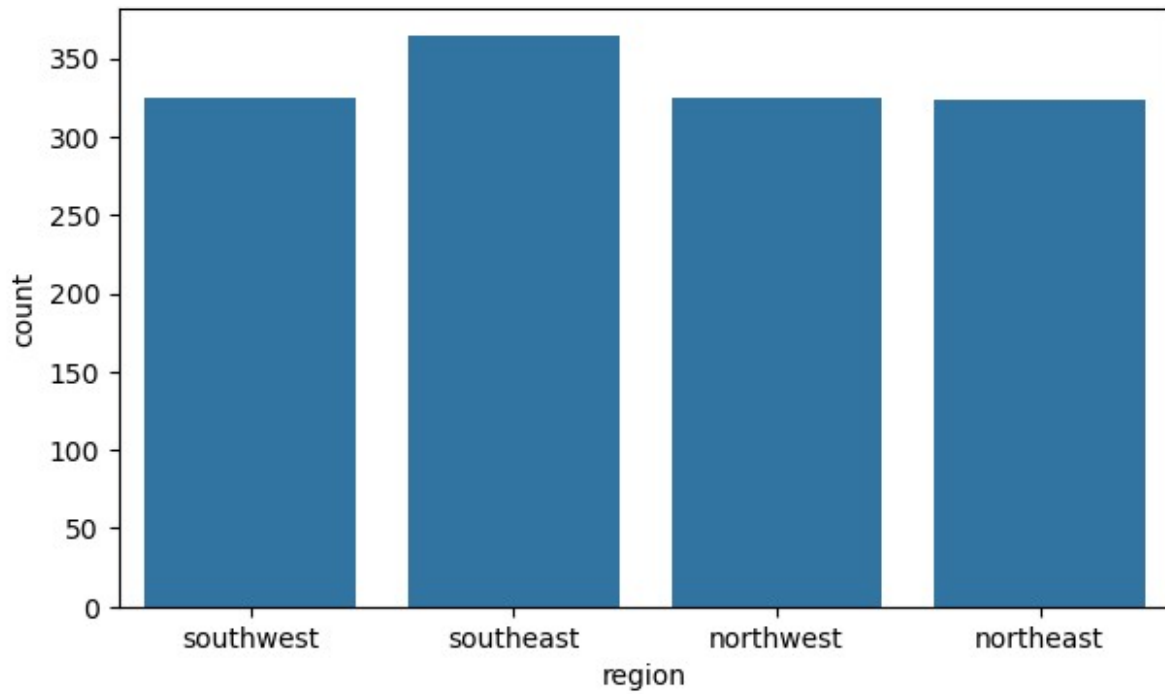
```
plt.figure(figsize=(7,4))  
sns.countplot(x='sex',data =df,color='black')  
plt.show()
```



```
plt.figure(figsize=(7,4))  
sns.countplot(x='smoker',data =df,color='green')  
plt.show()
```

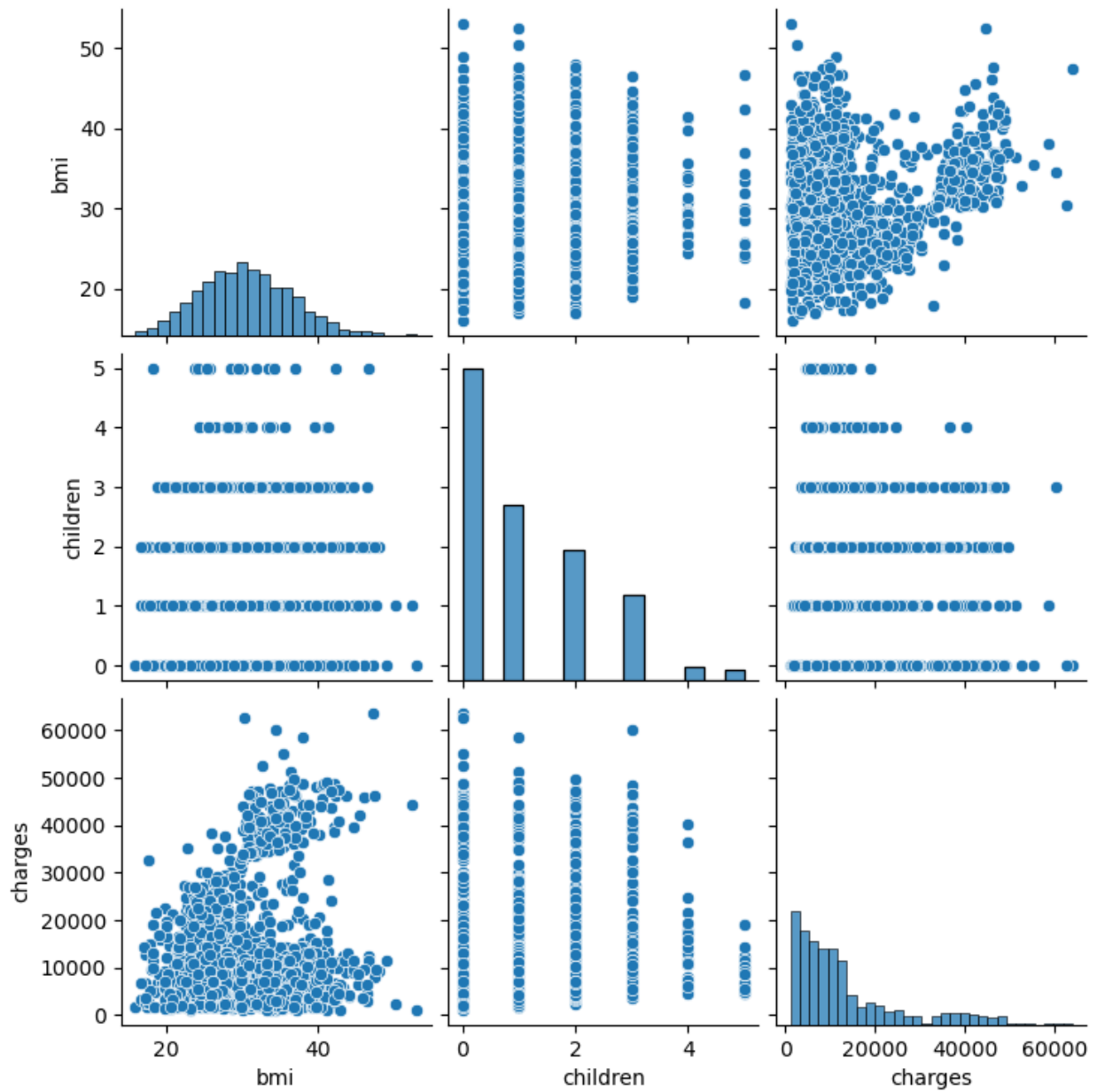


```
plt.figure(figsize=(7,4))  
sns.countplot(x='region',data =df)  
plt.show()
```

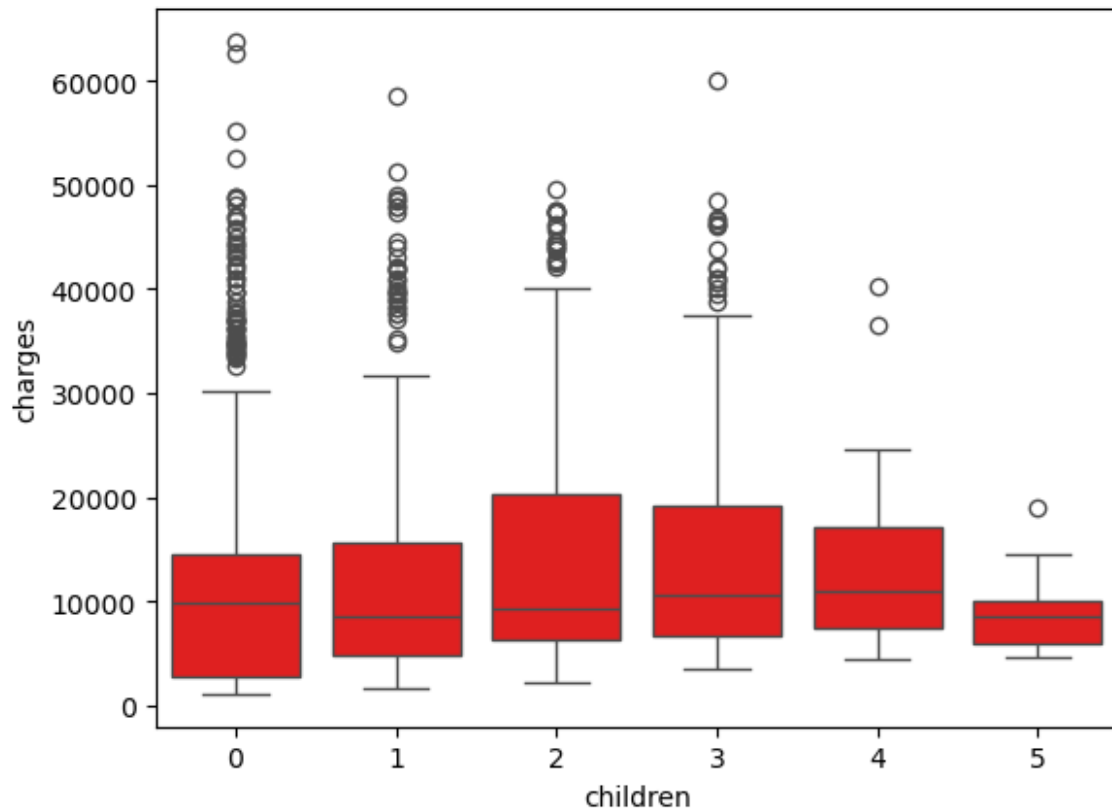


Made a pairplot to plot all possible interactions between a set of numeric variables.

```
sns.pairplot(data =df)  
plt.show()
```

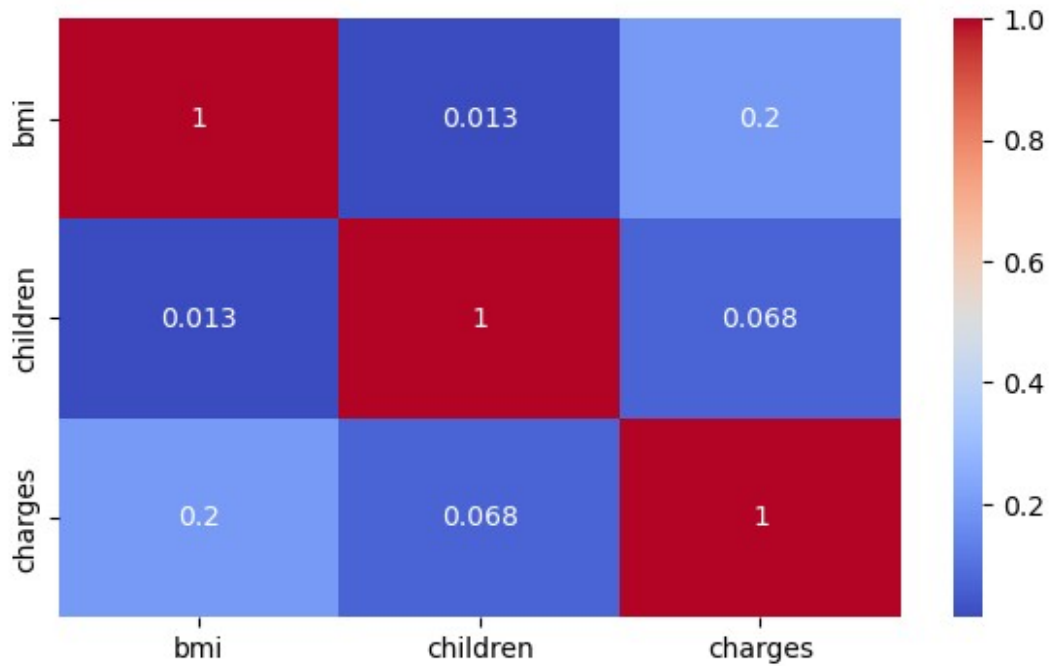


```
sns.boxplot(data = df, x= 'children',y='charges',color='red')
plt.show()
```



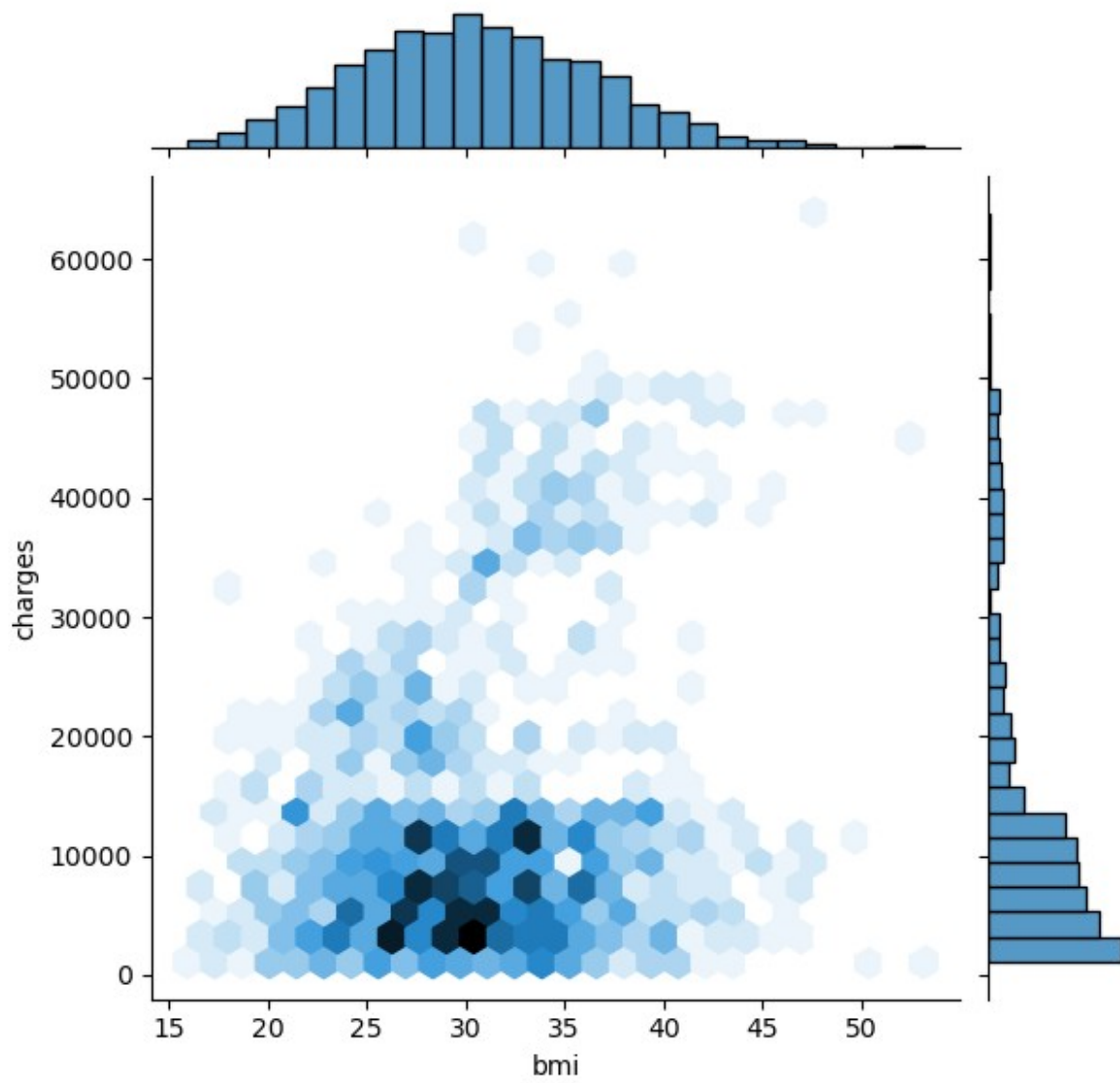
Created a heatmap to analyze the correlation values between the numeric variables.

```
df1 = df[['bmi', 'children', 'charges']]  
plt.figure(figsize=(7,4))  
sns.heatmap(df1.corr(),annot = True,cmap='coolwarm')  
plt.show()
```

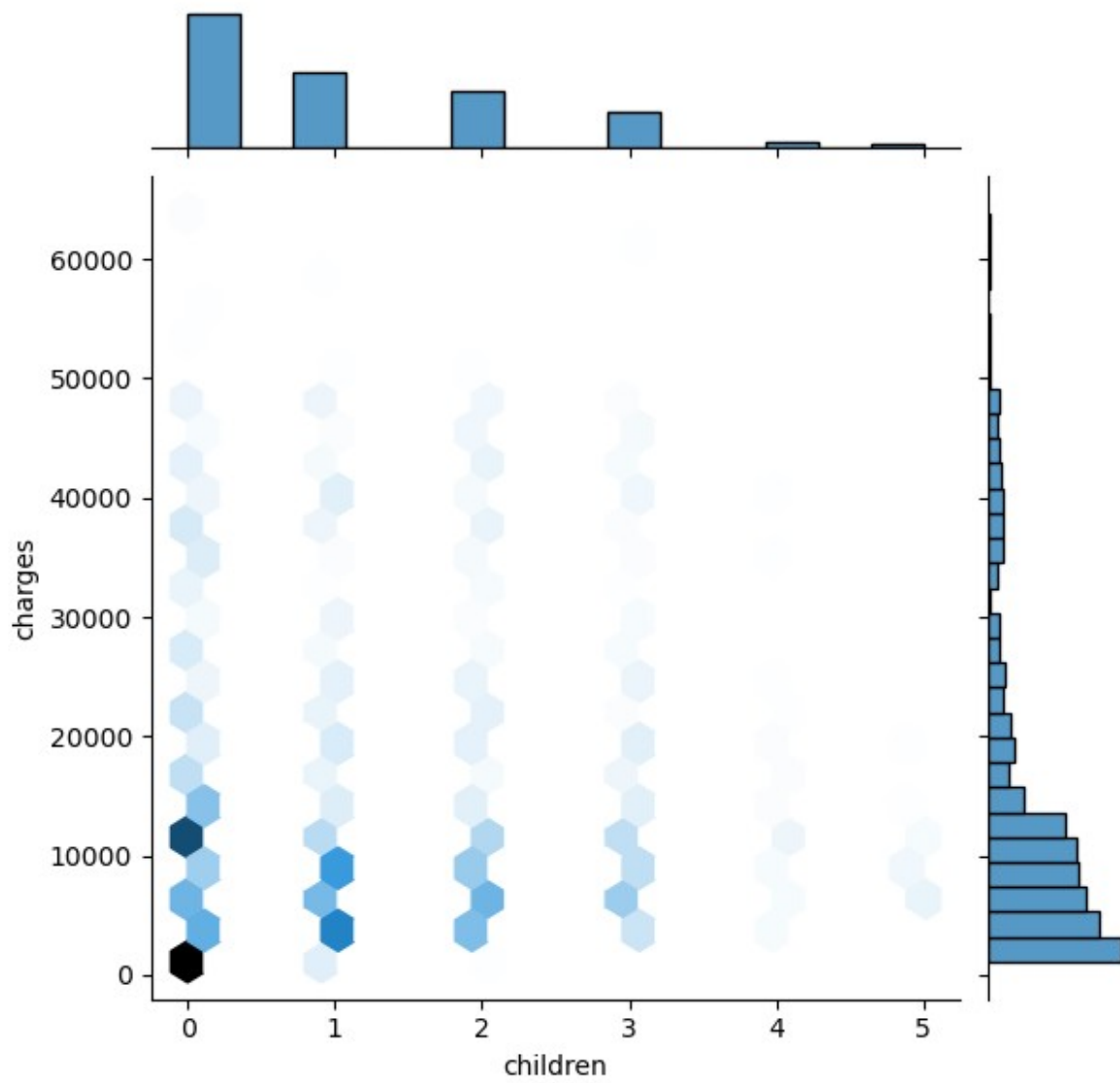
```
plt.figure(figsize=(7,4))  
sns.jointplot(data = df,x='bmi',y='charges',kind='hex')  
plt.show()
```

<Figure size 700x400 with 0 Axes>



```
plt.figure(figsize=(7,4))  
sns.jointplot(data = df,x='children',y='charges',kind='hex')  
plt.show()
```

<Figure size 700x400 with 0 Axes>



Checked if the number of premium charges for smokers or non-smokers is increasing as they are aging

```
plt.figure(figsize=(7,4))
sns.boxplot(data = df, y='age',x='smoker')
plt.show()
```

